1. Candy Lab
2. Descriptive Statistics: Descriptive Statistics Lab (edited: Teegarden)
3. Probability Topics: Probability Lab (edited: Teegarden)
4. Discrete Random Variables: Lab I (edited: Teegarden)
5. Continuous Random Variables: Lab I
6. Normal Distribution: Normal Distribution Lab I (edited: Teegarden)
7. Central Limit Theorem: Central Limit Theorem Lab I (edited: Teegarden)
8. Confidence Intervals: Confidence Interval Lab I (edited: Teegarden)
9. Confidence Intervals: Confidence Interval Lab II (edited: Teegarden)
10. Confidence Intervals: Confidence Interval Lab III (edited: Teegarden)
11. Hypothesis Testing of Single Mean and Single Proportion: Lab (Edited: Teegarden) 02
12. Hypothesis Testing of Two Means and Two Proportions: Lab I (edited: Teegarden 02)
13. The Chi-Square Distribution: Lab I (edited: Teegarden)
14. The Chi-Square Distribution: Lab II (edited: Teegarden)
15. Linear Regression and Correlation: Regression Lab II (edited: Teegarden)
16. Collaborative Statistics: Group Project - Teegarden

Candy Lab
This lab allows students to practice and demonstrate techniques used to generate systematic samples. Students will have the opportunity to create relative frequency tables and interpret results based on different data groupings. Labs modified to include minitab usage. This replaces the Chapter 1, Lab 1 from the Collaborative Statistics by Dean and Illowsky.

Candy Lab

Name:

Student Learning Outcomes

- The student will construct Relative Frequency Tables.
- The student will interpret results and their differences from different data groupings.
- The student will illustrate the data using pie charts and bar graphs.

General Directions

In class, answer the initial questions on the lines provided and complete the tables. The write-up questions should be answered in **paragraph form** and typed. The graphs must be generated in Minitab and may then be copied into the write-up or attached at the end of the lab. To save paper, please copy the graphs and paste them into Word so that more than one graph can be printed per page.

Data Collection

Before you open your candy, make you first prediction about the distribution of the colors of this candy. This prediction will be made with no knowledge about the color distribution except your personal experience. (There are no wrong answers.)

1. How many candies do you think there are in your package?

2. Which color do you think will occur the most often?

3. Which color do you think will occur the least?

Open your candy and sort them by color. DO NOT EAT ANY AT THIS TIME!

1. How many candies do you have?

2. Which color occurred the most often?

3. Which color occurred the least?

4. Based on your individual sample, what do you think the actual color distribution is for this candy? Predict a % for each color and explain your reasoning. You can only base your prediction on what you see in your sample, not what you know about the candy.

YOU MAY NOW EAT YOUR CANDY NOW

Summarizing the Data

Complete the three relative frequency tables below using your personal data, your group data and the class data.

Individual Bag Frequency Table

Color	Frequency	Relative Frequency

Group Color Frequency Table

Color	Frequency	Relative Frequency

Class Frequency Table

Color	Frequency	Relative Frequency

Graphs

1. Illustrate the data for each set (individual, group, class) by inputting the colors in C1, individual frequencies in C2, group frequencies in C3, and class frequencies in C4 and create a pie and bar chart for each.

2. You may copy the graphs into the write-up or attach them to your Lab. If you choose not to include the graphs in the body of the write-up, please copy and paste them into a Word document so as not to printout 6 pages of graphs.

Write – up

Answer the following questions in paragraph form. To compare/contrast the data you should refer to at least three key values. Either where the data is similar or where it is very different. Be sure to use the information from your charts and graphs to justify your statements using the data.

1. Using the tables and graphs, compare/contrast the results for **your data and the group's combined data**. Use at least three examples to justify your answer.

2. Using the tables and graphs, compare/contrast the results for **your data and the class' combined data**. Use at least three examples to justify your answer.

3. Using the tables and graphs, compare/contrast the results for **the group's combined data and the class' combined data.** Use at least three examples to justify your answer.

4. Which of the three data sets would you use to best predict the distribution of colors for this candy? Why? What would you predict the actual distribution for this candy may be?

Descriptive Statistics: Descriptive Statistics Lab (edited: Teegarden)
Labs changed to incorporate mini-tabs.

Descriptive Statistics Lab

Name:

A. Student Learning Objectives

• The student will construct a histogram and a box plot.

• The student will calculate univariate statistics.

• The student will examine the graphs to interpret what the data implies.

B. Collect the Data

Record the number of pairs of shoes you own:

1. Randomly survey 20 people. Record their values. Survey Results

_____ _____ _____ _____ _____ _____ _____ _____ _____ _____

_____ _____ _____ _____ _____ _____ _____ _____ _____ _____

2. Construct a histogram using Minitab. Choose an appropriate scale and use boundary values (cut points).

3. Calculate the following: Be sure to include the formulas and the appropriate values. Show your work

• =

• s =

4. Are the data discrete or continuous? How do you know? Use complete sentences.

5. Describe the shape of the histogram. Use 2 – 3 complete sentences.

C. Analyze the Data

1. Determine the following and show your work where appropriate:

• Minimum value =

• Median =

• Maximum value =

• First quartile =

• Third quartile =

• IQR =

2. Using Minitab, construct a box plot of data.

3. What does the shape of the box plot imply about the concentration of data? Use 2 – 3 complete sentences.

4. What does the IQR represent in this problem? (reference your values)

5. Are there any potential outliers? Which value(s) is (are) it (they)?

Use the formula to calculate the two end values used to determine if a data value is an outlier.

upper =

lower =

6. Show your work to find the value that is 1.5 standard deviations:

a. Above the mean:

b. Below the mean:

c. What percent of the data does Chebyshev's theorem state lies within 1.5 standard deviations of the mean? (show your work.)

d. What percentage of your data actually falls within 1.5 standard deviations of the mean? How does this compare to the value you calculated in part c above?

7. How does the standard deviation help you to determine concentration of the data and whether or not there are potential outliers?

Probability Topics: Probability Lab (edited: Teegarden)
This module presents students with a lab exercise allowing them to apply their understanding of Probability. Using a Minitab simulation, students will compare theoretical probabilities with experimental probabilities as it relates to dice. They will also investigate the effect of sample size.

Probability Lab

I. Student Learning Outcomes:

• The student will calculate theoretical and empirical probabilities.

• The student will appraise the differences between the two types of probabilities.

• The student will demonstrate an understanding of long-term probabilities.

II. Theoretical probability for the sum of two dice

Begin by looking at Theoretical probabilities for the sum of two dice. Let the value in the first row be the result for Die 1 and the value in the first column be the value for Die 2. Input the sum of the corresponding row and column in each box.

+	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Using the following table, record the theoretical probabilities.

Sum	2	3	4	5	6	7	8	9	10	11	12
Count											
Probability											

Using the table above, determine the following theoretical probabilities

1. P(sum less than 5) = _____

2. P(sum at least 9) = _____

3. P(sum at most 6) = _____

4. P(sum more than 7) = _____

5. P(sum between 3 and 8) = _____

6. P(sum less than 11) = _____

III. Experimental (empirical) probability for the sum of two dice

Rolling the dice Using Minitab, simulate rolling two dice 360 times and finding the sum. Use Calc → Random Data → Integer, 360 rows and save in die 1, die 2.

Then use the Calc → Row Statistics. Select Sum, the two die columns and save in Sum Use Stat → Tables → Tally to summarize the data.

Record the experimental probabilities:

Sum	2	3	4	5	6	7	8	9	10	11	12
Count											
Probability											

Using the table above, determine the following experimental probabilities

1. P(sum less than 5) = _____

2. P(sum at least 9) = _____

3. P(sum at most 6) = _____

4. P(sum more than 7) = _____

5. P(sum between 3 and 8) = _____

6. P(sum less than 11) = _____

IV. Essay Questions (On a separate sheet of paper, answer these questions in complete sentences.)

1. How do the empirical probabilities compare to the theoretical probabilities? (You may wish to convert the probabilities to percentages for ease of comparison.)

2. If you increased the number of times you rolled the dice to 720, would the empirical probability values change? Why?

Rerun the simulation and record your results.

Sum	2	3	4	5	6	7	8	9	10	11	12
Count											
Probability											

3. Did the increase in the number of trials cause the empirical probabilities and theoretical probabilities to be closer together or farther apart? Why? (You may wish to convert the probabilities to percentages for ease of comparison.)

ATTACH THE SESSION WINDOW WITH YOUR RESULTS AND THE ESSAY ANSWERS TO THIS COVER SHEET.

Discrete Random Variables: Lab I (edited: Teegarden)
This module allows students to explore concepts related to discrete random variables through the use of a simple playing card experiment. Students will compare empirical data to a theoretical distribution to determine if the experiment fist a discrete distribution. This lab involves the concept of long-term probabilities. Labs changed to incorporate mini-tabs.

Discrete Probability Lab

Name:

Student Learning Outcomes:

- The student will compare empirical data and a theoretical distribution to determine if everyday experiment fits a discrete distribution.
- The student will demonstrate an understanding of long-term probabilities.

Procedure: The experiment procedure is to pick one card from a deck of shuffled cards.

1. The theoretical probability of picking a diamond from a deck is:
2. Shuffle a deck of cards and pick one card from it and record whether it was a diamond or not a diamond.
3. Put the card back and reshuffle.
4. Do this a total of 10 times and record the number of diamonds picked.
5. What is the experimental probability of drawing a diamond?
6. How does the experimental probability compare to the theoretical probability? (high/low/about the same)

Using Minitab, simulate this experiment (drawing a card 10 times and recording the number of diamonds) for a total of 50 times. **Use Calc -> Random data -> Binomia**l.

I Organize the Data:

Summarize the data generated in Minitab and include determine both the frequency and relative frequency. Record the result here:

X	Frequency	Relative Frequency
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

2. Calculate the following using Minitab. (include the session window)

$\bar{x}$= _____ s = _____

3. Construct a bar chart of the experimental data using the relative frequency as the vertical axis and attach it to this cover sheet. Don’t forget a title and labels for the graph

II. Theoretical Distribution

1. Using Minitab, build the theoretical PDF chart for X based on the distribution in the section above.

X	P(X)
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

2. Calculate the following, **indicating the formulas**:

μ = _____ σ = _____

3. Constuct a graph of the theoretical distribution by using:

graph → probability distribution plot → single view → Binomial

Attach the graph to this cover sheet.

III. Using the Data

Using the Theoretical probability table generated by Minitab, determine the following theoretical probabilities, rounding to 4 decimal places:

P(X = 3) =_____ P(2 < X < 5) = _____ P(X > 8) _____

Using the data from the Minitab simulation, determine the following empirical (experimental) probabilities:

P(X = 3) = _____ P(2 < X < 5) = _____ P(X > 8) _____

IV. Discussion Questions:

Answer the following in complete sentences on a separate sheet of paper and attach it to this cover sheet.

1. Knowing that data vary, describe **two similarities between the graphs and distributions** of the theoretical and experimental distributions.
2. Describe the **two most significant differences between the graphs or distributions** of the theoretical and experimental distributions.
3. Suppose that the experiment had been repeated 500 times. Would you expect the frequency table and bar chart in part I above to change?

How and Why? Repeat the experiment and justify your answer. (Be sure to include the data summary and bar chart.)

Continuous Random Variables: Lab I
In this lab exercise, students will compare and contrast empirical data using Minitab with the Uniform Distribution. Note: This module is based on a student being able to access the Minitab statistical program. This modu

Continuous Distribution Lab

Name:

I - Student Learning Outcomes:

• The student will compare and contrast empirical data from a random number generator with the Uniform Distribution.

II - Theoretical Distribution

The theoretical distribution of X is X~U (0, 1). Use it for this part. In theory,

μ = _____ σ = _____ 1st quartile = _____

40th percentile = _____ 3rd quartile = _____Median = _____

III Collect the Data

Use Minitab to generate 100 values between 0 and 1 (inclusive). (Calc → Random Data → Uniform) Using Minitab, calculate the following (include the session window):

= _____ s = _____ 1st quartile = _____

40th percentile = _____ (justify) 3rd quartile = _____ median = _____

IV - Comparing the Data

1. For each part below, use a complete sentence to comment on how the value obtained from the experimental data (see part III) compares to the theoretical value you expected from the distribution in section II. (**How it is reflected in the corresponding data. Be specific!**)

a. minimum value:

b. first quartile:

c. median:

d. third quartile

e. maximum value:

f. width of IQR:

V - Plotting the Data and Interpreting the Graphs.

1. What does the probability graph for the theoretical distribution look like? Draw it here and label the axis.

2. Use Minitab to construct a histogram a using 5 bars and density as the y-axis value. Be sure to attach the graphs to this lab.

a. Describe the shape of the histogram. Use 2 - 3 complete sentences. (Keep it simple. Does the graph go straight across, does it have a V shape, does it have a hump in the middle or at either end, etc.? One way to help you determine the shape is to roughly draw a smooth curve through the top of the bars.)

b. How does this histogram compare to the graph of the theoretical uniform distribution? Draw the horizontal line which represents the theoretical distribution on the histogram for ease of comparison. Be sure to use 2 – 3 complete sentences.

3. Draw the box plot for the theoretical distribution and label the axis.

4. Construct a box plot of the experimental data using Minitab and attach the graph.

a. Do you notice any potential outliers? _____

If so, which values are they? _____

b. Numerically justify your answer using the appropriate formulas.

c. How does this plot compare to the box plot of the theoretical uniform distribution? Be sure to use 2 – 3 complete sentences.

VI - Increasing the sample size. Repeat the simulation with 500 data values.

1. Using Minitab, calculate the following (include the session window):

= _____ s = _____ 1^{st} quartile = _____

40th percentile = _____ (justify) 3rd quartile = _____ median = _____

2. Does this data appear to reflect the theoretical data more closely than the original? Be sure to use 2 – 3 complete sentences. (**Be specific.**)

3. Create a histogram with 5 bars and using density for the y-axis and box plot for this data. (attach to this lab)

4. How do these compare to the theoretical distribution? Be sure to use 2 – 3 complete sentences. (**Be specific.**)

Normal Distribution: Normal Distribution Lab I (edited: Teegarden)
Labs changed to incorporate mini-tabs.

Normal Distribution Lab

Name:

I Student Learning Outcome:

* The student will compare and contrast empirical data and a theoretical distribution.

* Find Probabilities for specific Normal Distributions

II The Situation

It is generally accepted that the mean body temperature is 98.6 degrees. If a sample of size 100 resulted in a sample mean of 98.3 degrees with a standard deviation of 0.64 degrees. Does this sample suggest that the mean body temperature is actually lower than 98.6 degrees?

III Simulation: To answer the question, complete the following simulation.

Using Minitab (Calc -> Random Data-> Normal), generate 100 values from a normally distributed population with a mean of 98.6 degrees and a standard deviation of 0.64 degrees (using the sample standard deviation given in the situation since the population deviation is unknown). Repeat the simulation 9 more times for a total of 10. (Requesting the data be stored in c2-c10 will generate the remaining 9 columns of data with one command.)

IV Data Collection

Use Stats -> Basic Stats -> Display Descriptive and select all 10 columns to determine the sample mean for each data set. Record the values below and

include the session window with this lab.

x_1 = _____ x_2 = _____ x_3 = _____ x_4 = _____

x_5 = _____ x_6 = _____ x_7 = _____ x_8 = _____

x_9 = _____ x_{10} = _____

V Analyze the Data – Using complete sentences.

Based on your simulation, do you think that a sample of size 100 with a mean temperature of 98.3 is reasonable? Answer using 2 – 3 complete sentences.

VI Finding Probabilities for Normal Distribution

For each of the following, first write the question in **symbolic form** and then using Minitab (Calc -> Probability Distributions -> Normal), find the probabilities. (attach your session window to this lab)_____

1. Given a population with a normal distribution, a mean of 0, and a standard deviation of 1, find the probability of a value less than 1.25 _____

2. Given a population with a normal distribution, a mean of 25, and a standard deviation of 3, find the probability of a value greater than 21.25._____

3. Given a population with a normal distribution, a mean of 100, and a standard deviation of 20, find the probability of a value between 87 and 122._____

4. Given a population with a normal distribution, a mean of 150, and a standard deviation of 35, what value has an area of 0.34 to the left? _____

5. Given a population with a normal distribution, a mean of 150, and a standard deviation of 35, what value has an area of 0.34 to the right?

_____

6. Given a population with a normal distribution, a mean of 15, and a standard deviation of 2, what value has an area of 0.8 to the left?

_____

7. Given a population with a normal distribution, a mean of 200, and a standard deviation of 15, which two values form the upper and lower boundary of the middle 80%?_____

Central Limit Theorem: Central Limit Theorem Lab I (edited: Teegarden)

Class Time:

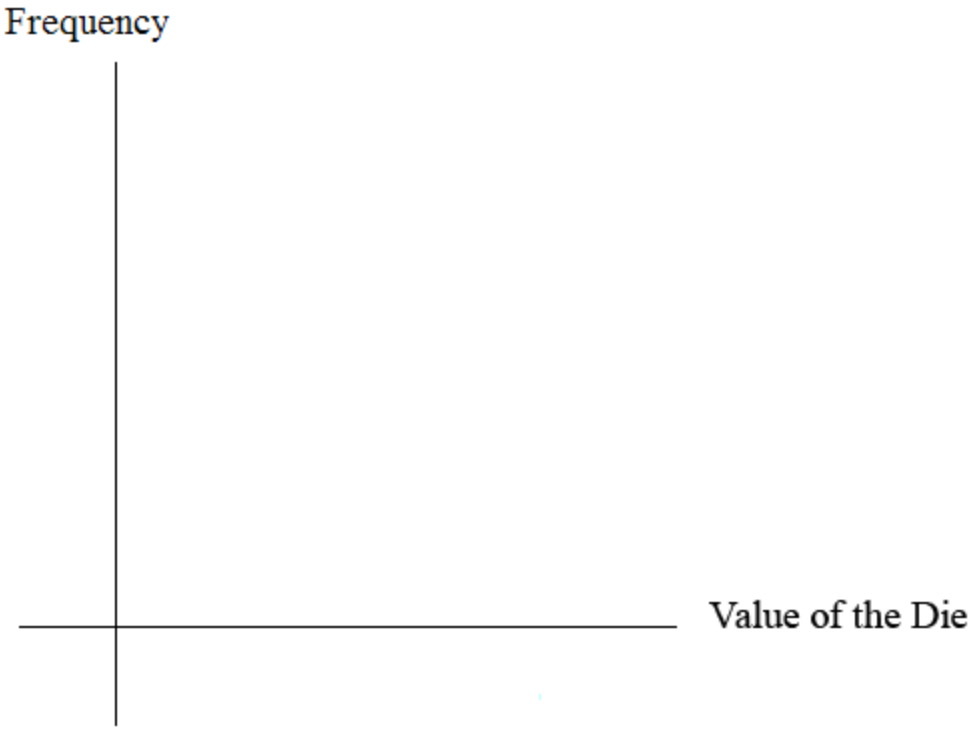
Name:

Student Learning Outcomes:

- The student will examine properties of the Central Limit Theorem.

Collect the Data

1. Using the random number generator in minitab, simulate the tossing of a single die 60 times. Calc -> Random Data -> Integer
2. Using Stat -> Tables -> Tally, summarize the data
3. Construct a histogram using Minitab and then sketch the graph using a ruler and pencil. Scale the axes.

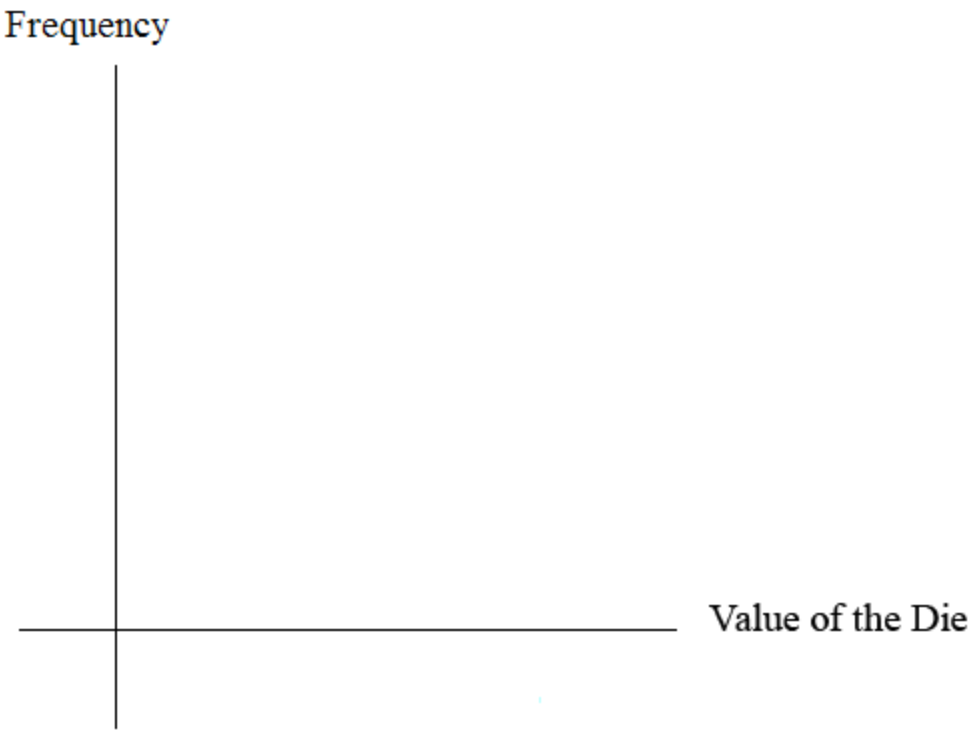


4. Caluclate the following:
    - **a** x =
    - **b** s =
    - **c** $n = 1$ (single die)
5. Draw a smooth curve through the tops of the bars of the histogram. Use 1 – 2 complete sentences to describe the general shape of the curve.

Collecting Averages of Pairs

Repeat steps 1 - 5 (of the section above titled "Collect the Data") with one exception. Instead of recording the value of a single die, record the average of two dice. Use Minitab and generate 50 rows with two columns. Then use the Calc -> Row Statistics and select mean. Then use Stats -> Tables -> Tally to summarize the data.

1. Construct a histogram. Scale the axes using the same scaling you did for the section titled "Collecting the Data". Sketch the graph using a ruler and a pencil.



2. Calculate the following:
    - **a** $\bar{x}$ =
    - **b** s =
    - **c** $n = 2$ (surveying one person at a time)

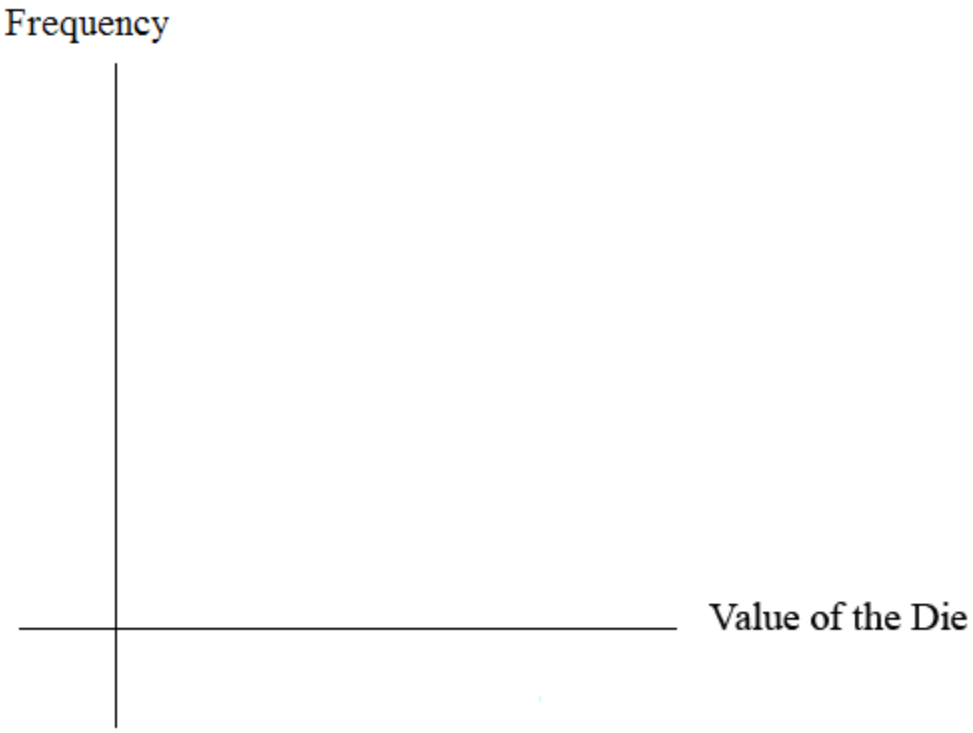
3. Draw a smooth curve through tops of the bars of the histogram. Use 1 – 2 complete sentences to describe the general shape of the curve.

Collecting Averages of Groups of Five

Repeat steps 1 – 5 (of part I) with one exception. Instead of recording the value for a single die, record the average value for each of the 50 groups of 5 die tosses.

1. Generate fifty **groups of 5** die tosses. Record the values of the average of their value.

2. Construct a histogram. Scale the axes using the same scaling you did for section titled "Collect the Data". Sketch the graph using a ruler and a pencil.



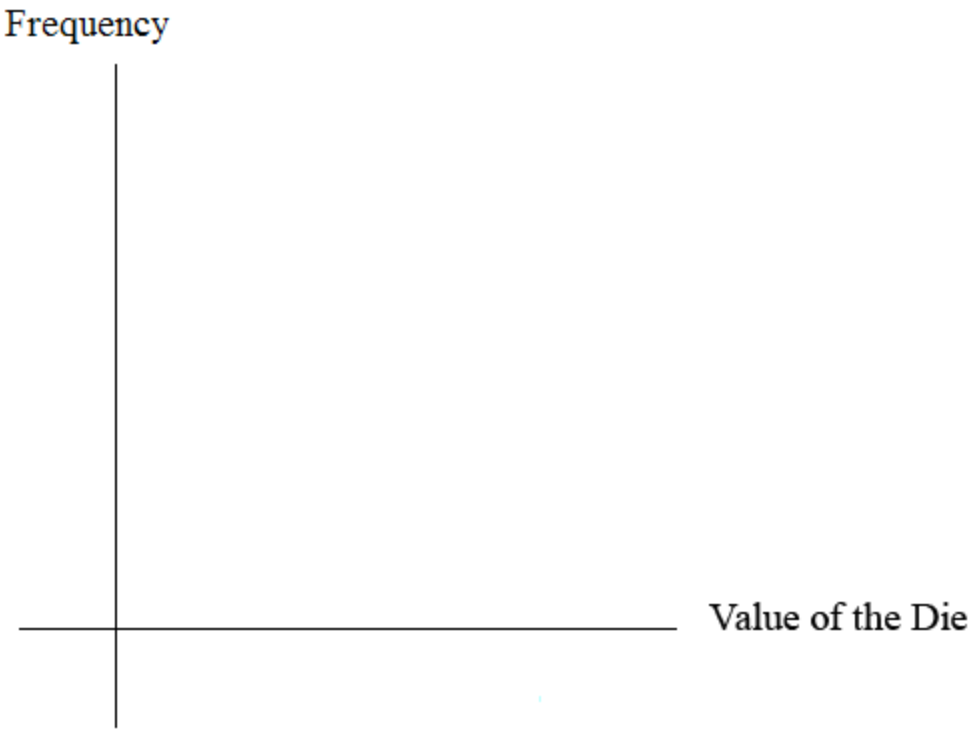
3. Calculate the following:
    - **a** x =
    - **b** s =
    - **c** $n = 5$ (surveying five people at a time)

4. Draw a smooth curve through tops of the bars of the histogram. Use 1 – 2 complete sentences to describe the general shape of the curve.

Collecting Averages of Groups of 20

Repeat steps 1 – 5 (of part I) recording the average value for each of the 50 groups of 20 die tosses.

1. Generate fifty **groups of 20** die tosses. Record the values of the average of their value.
2. Construct a histogram. Scale the axes using the same scaling you did for section titled "Collect the Data". Sketch the graph using a ruler and a pencil.



3. Calculate the following
4. Draw a smooth curve through tops of the bars of the histogram. Use 1 – 2 complete sentences to describe the general shape of the curve.

Discussion Questions

1. As n changed, why did the shape of the distribution of the data change? Use 1 – 2 complete sentences to explain what happened.
2. In the section titled "Collect the Data", what was the approximate distribution of the data? X ~
3. In the section titled "Collecting Averages of Groups of Five", what was the approximate distribution of the data? X ~

4. In the section titled "Collecting Averages of Groups of Twenty", what was the approximate distribution of the data? X ~
5. In 1 – 2 complete sentences, explain any differences in your answers to previous three questions.

Confidence Intervals: Confidence Interval Lab I (edited: Teegarden)

Class Time:

Name:

Student Learning Outcomes:

- The student will calculate the 90% confidence interval for the average cost of a home in the area in which this school is located.
- The student will interpret confidence intervals.
- The student will examine the effects that changing conditions has on the confidence interval.

Collect the Data

Check the Real Estate section in your local newspaper or website. (Note: many papers only list them one day per week. Also, we will assume that homes come up for sale randomly.) Record the sales prices for 35 randomly selected homes recently listed in the county.

1. Complete the table:

_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____

Describe the Data

1. Compute the following:

- **a** =
- **b** =
- **c** =

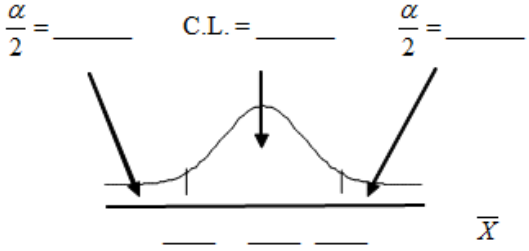
2. Define the Random Variable , in words. =
3. State the estimated distribution to use. Use both words and symbols.

Find the Confidence Interval

1. Calculate the confidence interval and the error bound.

   - **a**Confidence Interval:
   - **b**Error Bound:

2. How much area is in both tails (combined)? =
3. How much area is in each tail? — =
4. Fill in the blanks on the graph with the area in each section. Then, fill in the number line with the upper and lower limits of the confidence interval and the sample mean.



5. Some students think that a 90% confidence interval contains 90% of the data. Use the list of data on the first page and count how many of the data values lie within the confidence interval. What percent is this? Is this percent close to 90%? Explain why this percent should or should not be close to 90%.
6. How many house prices would be needed in the sample to ensure that the error was no more than $2000 for the 90% confidence interval?

Describe the Confidence Interval

1. In two to three complete sentences, explain what a Confidence Interval means (in general), as if you were talking to someone who has not taken statistics.
2. In one to two complete sentences, explain what this Confidence Interval means for this particular study.

Use the Data to Construct Confidence Intervals

1. Using the above information, construct a confidence interval for each confidence level given.

Confidence level	EBM / Error Bound	Confidence Interval
50%		
80%		
95%		
99%		

2. What happens to the EBM as the confidence level increases? Does the width of the confidence interval increase or decrease? Explain why this happens.

Effect of an Outlier

Suppose one of the values input incorrectly. Choose one data value and increase the amount by adding two extra zeroes.

1. Calculate the 90% confidence interval: _____
2. Calculate the Error Bound: _____
3. How does the outlier effect the width of the confidence interval?

Confidence Intervals: Confidence Interval Lab II (edited: Teegarden)

Class Time:

Name:

Student Learning Outcomes:

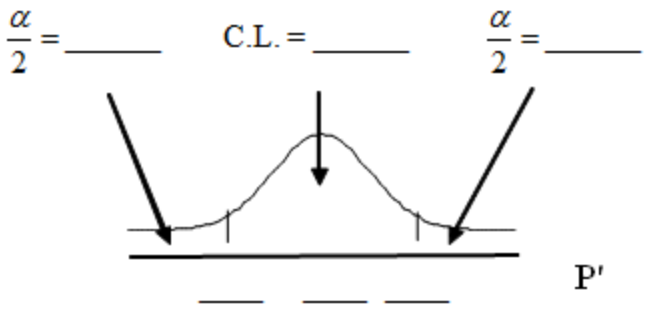
- The student will calculate the 90% confidence interval for proportion of students in this school that were born in this state.
- The student will interpret confidence intervals.
- The student will examine the effects that changing conditions has on the confidence interval.

Collect the Data

1. Survey the students in your class, asking them if they were born in this state. Let X = the number that were born in this state.
    - **a** n =_____
    - **b** x =_____
2. Define the Random Variable P' in words.
3. State the estimated distribution to use.

Find the Confidence Interval and Error Bound

1. Calculate the confidence interval and the error bound.
    - **a** Confidence Interval:
    - **b**Error Bound:
2. How much area is in both tails (combined)? α=
3. How much area is in each tail? $\frac{\alpha}{2}$ =
4. Fill in the blanks on the graph with the area in each section. Then, fill in the number line with the upper and lower limits of the confidence interval and the sample proportion.



5. How large a sample would be needed to ensure that the error for the 90% confidence interval is only 5%? Use the previous data as an estimate for p.
6. Recalculate the sample size assuming there is no previous data.

Describe the Confidence Interval

1. In two to three complete sentences, explain what a Confidence Interval means (in general), as if you were talking to someone who has not taken statistics.
2. In one to two complete sentences, explain what this Confidence Interval means for this particular study.
3. Using the above information, construct a confidence interval for each given confidence level given.

Confidence level	EBP / Error Bound	Confidence Interval
50%		
80%		
95%		

Confidence level	EBP / Error Bound	Confidence Interval
99%		

4. What happens to the EBP as the confidence level increases? Does the width of the confidence interval increase or decrease? Explain why this happens.

Confidence Intervals: Confidence Interval Lab III (edited: Teegarden)

Class Time:

Name:

Student Learning Outcomes:

- The student will calculate a 90% confidence interval using the given data.
- The student will examine the relationship between the confidence level and the percent of constructed intervals that contain the population average.

Given:

1.

59.4	71.6	69.3	65.0	62.9
66.5	61.7	55.2	67.5	67.2
63.8	62.9	63.0	63.9	68.7
65.5	61.9	69.6	58.7	63.4
61.8	60.6	69.8	60.0	64.9
66.1	66.8	60.6	65.6	63.8
61.3	59.2	64.1	59.3	64.9
62.4	63.5	60.9	63.3	66.3
61.5	64.3	62.9	60.6	63.8
58.8	64.9	65.7	62.5	70.9
62.9	63.1	62.2	58.7	64.7
66.0	60.5	64.7	65.4	60.2
65.0	64.1	61.1	65.3	64.6

59.2	61.4	62.0	63.5	61.4
65.5	62.3	65.5	64.7	58.8
66.1	64.9	66.9	57.9	69.8
58.5	63.4	69.2	65.9	62.2
60.0	58.1	62.5	62.4	59.1
66.4	61.2	60.4	58.7	66.7
67.5	63.2	56.6	67.7	62.5

Heights of 100 Women (in Inches)

Listed above are the heights of 100 women. Use MiniTab to randomly select 10 data values.

2. Calculate the sample mean and sample standard deviation. Assume that the population standard deviation is known to be 3.3. With these values, construct a 90% confidence interval for your sample of 10 values. Write the confidence interval you obtained in the first space of the table below.
3. Now write your confidence interval on the board. As others in the class write their confidence intervals on the board, copy them into the table below:

_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____

90% Confidence Intervals

Discussion Questions

1. The actual population mean for the 100 heights given above is $\mu = 63.4$. Using the class listing of confidence intervals, count how many of them contain the population mean μ; i.e., for how many intervals does the value of μ lie between the endpoints of the confidence interval?
2. Divide this number by the total number of confidence intervals generated by the class to determine the percent of confidence intervals that contain the mean μ. Write this percent below.
3. Is the percent of confidence intervals that contain the population mean μ close to 90%?
4. Suppose we had generated 100 confidence intervals. What do you think would happen to the percent of confidence intervals that contained the population mean?
5. When we construct a 90% confidence interval, we say that we are **90% confident that the true population mean lies within the confidence interval.** Using complete sentences, explain what we mean by this phrase.
6. Some students think that a 90% confidence interval contains 90% of the data. Use the list of data given on the first page and count how many of the data values lie within the confidence interval that you generated on that page. How many of the 100 data values lie within your confidence interval? What percent is this? Is this percent close to 90%?
7. Explain why it does not make sense to count data values that lie in a confidence interval. Think about the random variable that is being used in the problem.
8. Suppose you obtained the heights of 10 women and calculated a confidence interval from this information. Without knowing the population mean μ, would you have any way of knowing **for certain** if your interval actually contained the value of μ? Explain.

Note:*This lab was designed and contributed by Diane Mathios.*

Hypothesis Testing of Single Mean and Single Proportion: Lab (Edited: Teegarden) 02

Class Time:

Name:

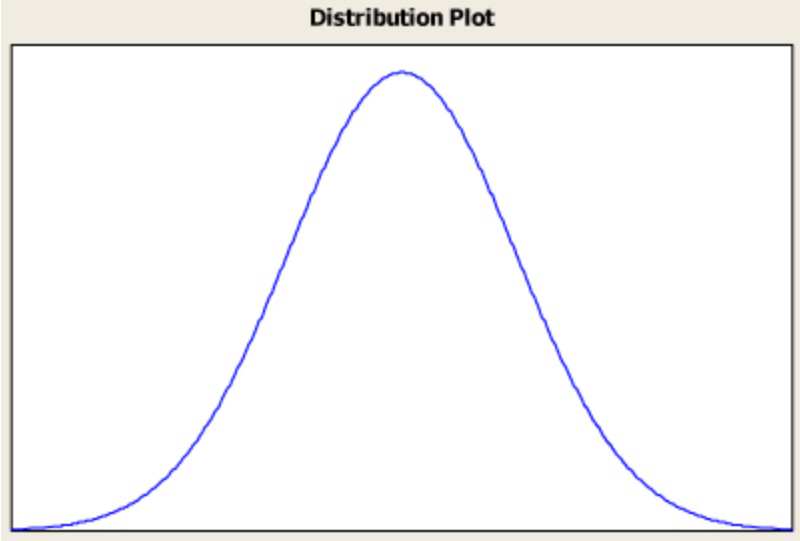
Student Learning Outcomes:

- The student will select the appropriate distributions to use in each case.
- The student will conduct hypothesis tests and interpret the results.

Television Survey

The data in the Testbook.mtw file lists the cost of 62 books required for classes in Summer 2008 at Mesa College. Students believe that they are spending on average $100 for their textbooks. Using the data for new books as the sample, conduct a hypothesis test to determine if the average cost of new textbooks at Mesa is lower.

1. H_o:
2. H_a:
3. In words, define the random variable. _____ =
4. The distribution to use for the test is:
5. Calculate the test statistic using your data.
6. Draw a graph and label it appropriately.
    - **a**Graph:



- **b**Calculate the p-value:

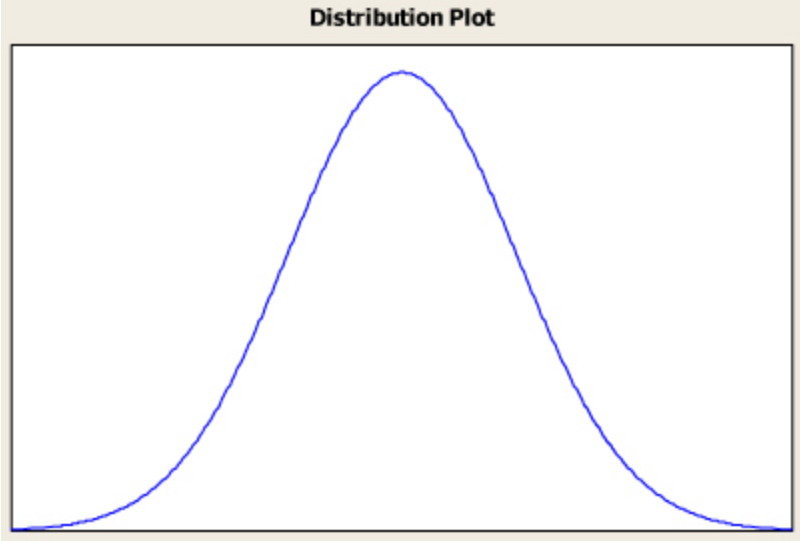
7. Do you or do you not reject the null hypothesis? Why?
8. Write a clear conclusion using a complete sentence.

Language Survey

According to the 2000 Census, about 39.5% of Californians and 17.9% of all Americans speak a language other than English at home. Using your class as the sample, conduct a hypothesis test to determine if the percent of the students at your school that speak a language other than English at home is different from 39.5%.

1. H_o:
2. H_a:
3. In words, define the random variable. _____ =
4. The distribution to use for the test is:
5. Calculate the test statistic using your data.
6. Draw a graph and label it appropriately. Shade the actual level of significance.

    - **a** Graph:



- **b**Calculate the p-value:

7. Do you or do you not reject the null hypothesis? Why?
8. Write a clear conclusion using a complete sentence.

Hypothesis Testing of Two Means and Two Proportions: Lab I (edited: Teegarden 02)

Class Time:

Name:

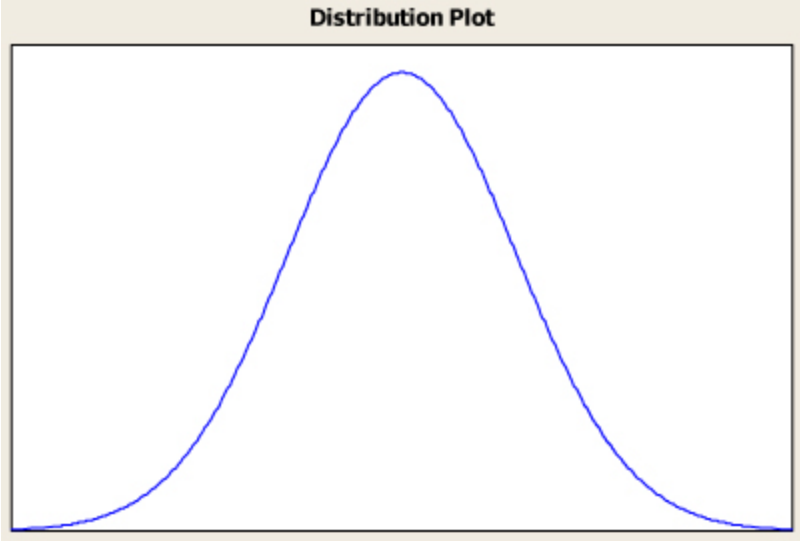
Student Learning Outcomes:

- The student will select the appropriate distributions to use in each case.
- The student will conduct hypothesis tests and interpret the results.

Increasing Stocks Survey

Look at yesterday's newspaper business section. Conduct a hypothesis test to determine if the proportion of New York Stock Exchange (NYSE) stocks that increased is greater than the proportion of NASDAQ stocks that increased. As randomly as possible, choose 40 NYSE stocks and 32 NASDAQ stocks and input the data into a Minitab worksheet. Complete the following statements.

1. H_o
2. H_a
3. In words, define the Random Variable. _____=
4. The distribution to use for the test is:
5. Calculate the test statistic using your data.
6. Draw a graph and label it appropriately. Shade the actual level of significance.
    - **a**Graph:



- **b**Calculate the p-value:

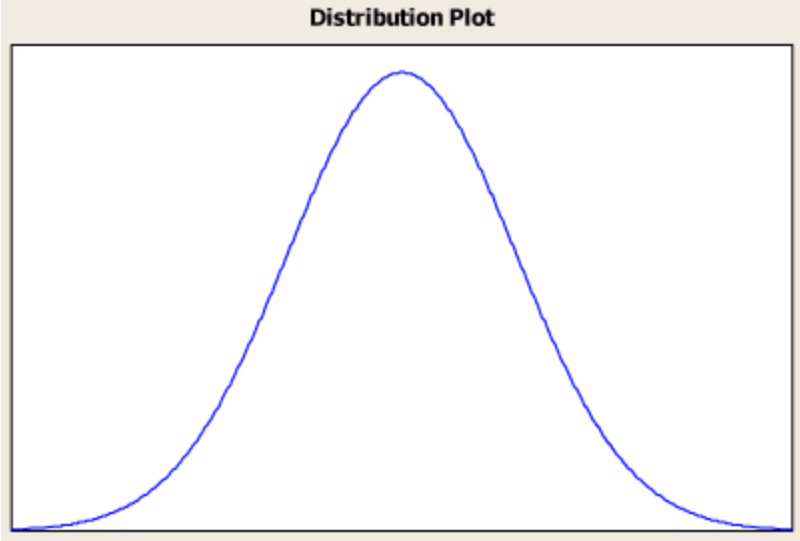
7. Do you reject or not reject the null hypothesis? Why?
8. Write a clear conclusion using a complete sentence.

Textbook Prices

The data in Textbook.mtw shows the price for both the new and used textbook price for books required for summer 2008 classes at Mesa College. Is it worthwhile to buy used textbooks?

1. H_o
2. H_a
3. In words, define the Random Variable. _____=
4. The distribution to use for the test is:
5. Calculate the test statistic using your data.
6. Draw a graph and label it appropriately. Shade the actual level of significance.

    - **a**Graph:



- **b**Calculate the p-value:

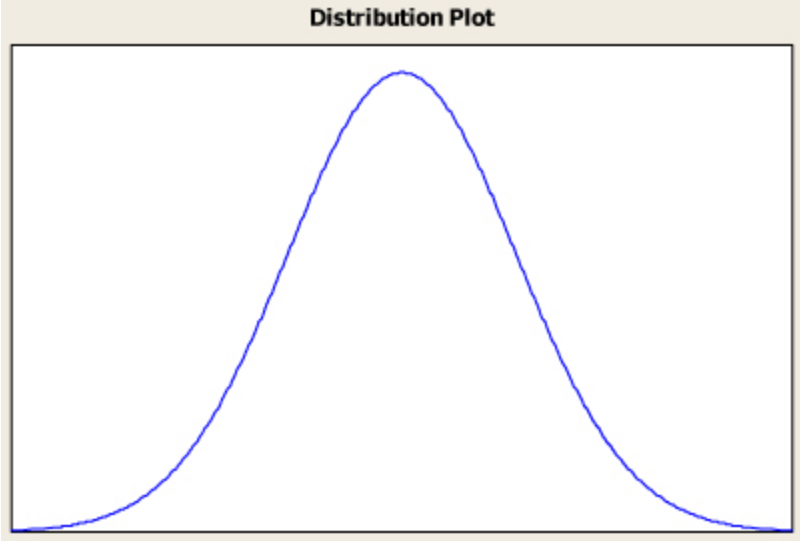
7. Do you reject or not reject the null hypothesis? Why?
8. Write a clear conclusion using a complete sentence.

Shoe Survey

Test whether women have, on average, more pairs of shoes than men. Include all forms of sneakers, shoes, sandals, and boots. Use your class as the sample.

1. H_o
2. H_a
3. In words, define the Random Variable. _____=
4. The distribution to use for the test is:
5. Calculate the test statistic using your data.
6. Draw a graph and label it appropriately. Shade the actual level of significance.

    - **a**Graph:



- **b**Calculate the p-value:

7. Do you reject or not reject the null hypothesis? Why?
8. Write a clear conclusion using a complete sentence.

The Chi-Square Distribution: Lab I (edited: Teegarden)

Class Time:

Name:

Student Learning Outcome:

- The student will evaluate data collected to determine if they fit either a uniform distribution.

Collect the Data

Three car-pooling students claimed that they missed their statistics test because they had a flat tire. On the make-up test the instructor asked them to identify the particular tire that went flat. The instructor assumed that the distribution would be uniform. To test this assumption,survey the class to determine the number the which tire they would select.

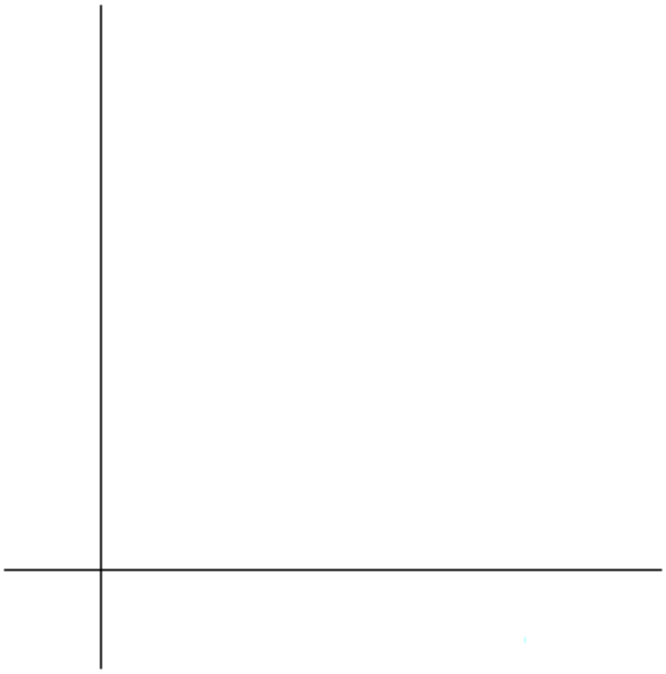
Tire	Left Front	Left Rear	Right Front	Right Rear
Number Selected				

Hypothesis Test

Conduct a hypothesis test to determine if the selection of a tire fits a Uniform Distribution.

1. :

2. :
3. Calculate the test statistic.
4. Find the p-value.
5. Sketch a graph of the situation. Label and scale the x-axis. Shade the area corresponding to the p-value.



6. State your decision.
7. State your conclusion in a complete sentence.
8. If in fact the students did not have a flat tire, do you think they will be caught out? Explain.

The Chi-Square Distribution: Lab II (edited: Teegarden)

Class Time:

Name:

Student Learning Outcome:

- The student will evaluate if there is a significant relationship between favorite type of snack and gender.

Collect the Data

1. Using your class as a sample, complete the following chart.

	sweets (candy & baked goods)	**ice cream**	**chips & pretzels**	**fruits & vegetables**	**Total**
male					
female					
Total					

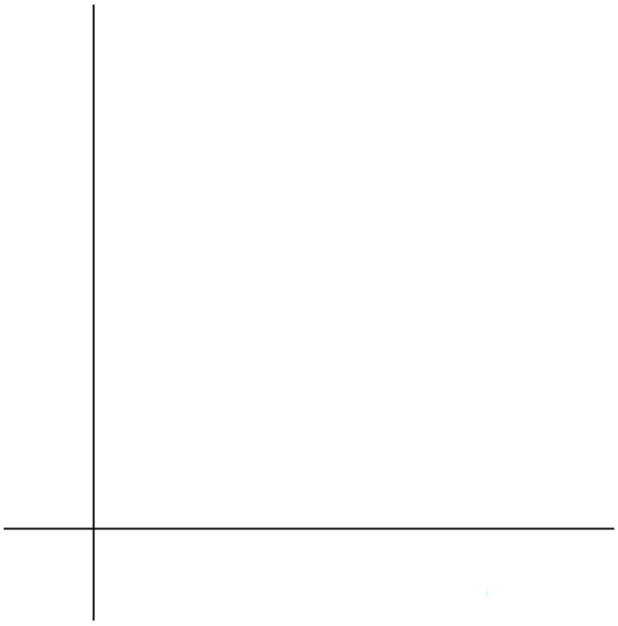
Favorite type of snack

2. Looking at the above chart, does it appear to you that there is dependence between gender and favorite type of snack food? Why or why not?

Determine the Classification

Conduct a hypothesis test to determine if the factors are independent

1. :
2. :
3. What distribution should you use for a hypothesis test?
4. Why did you choose this distribution?
5. Calculate the test statistic.
6. Find the p-value.
7. Sketch a graph of the situation. Label and scale the x-axis. Shade the area corresponding to the p-value.



8. State your decision.
9. State your conclusion in a complete sentence.

Discussion Questions

1. Is the conclusion of your study the same as or different from your answer to (I2) above?
2. Why do you think that occurred?

Linear Regression and Correlation: Regression Lab II (edited: Teegarden)

Class Time:

Name:

Student Learning Outcomes:

- The student will calculate and construct the line of best fit between two variables.
- The student will evaluate the relationship between two variables to determine if that relationship is significant.

Collect the Data

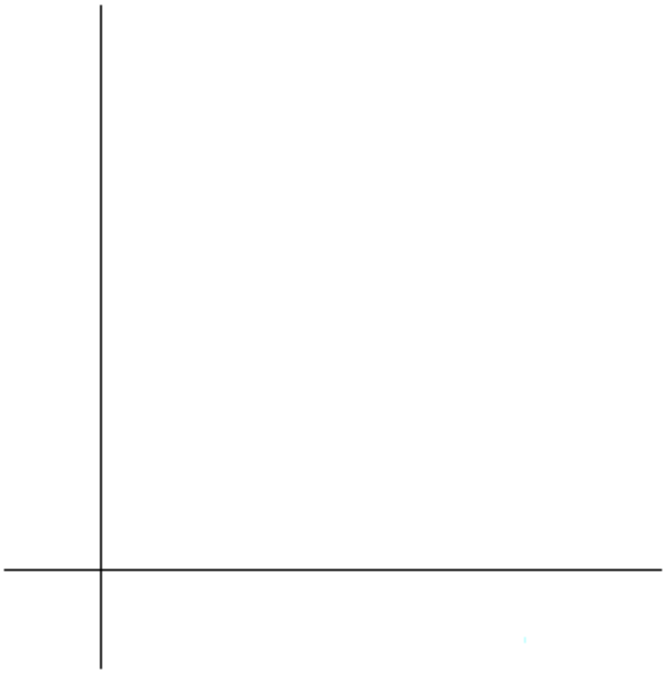
Survey 10 textbooks. Collect bivariate data (number of pages in a textbook, the cost of the textbook).

1. Complete the table.

Number of pages	Cost of textbook

2. Which variable should be the dependent variable and which should be the independent variable? Why?

3. Graph “number of pages” vs. “cost.” Plot the points on the graph. Label both axes with words. Scale both axes.



Does there appear to be a relationship between the size of the book and it's cost?

Analyze the Data

Enter your data into your MiniTab. Record the following information.

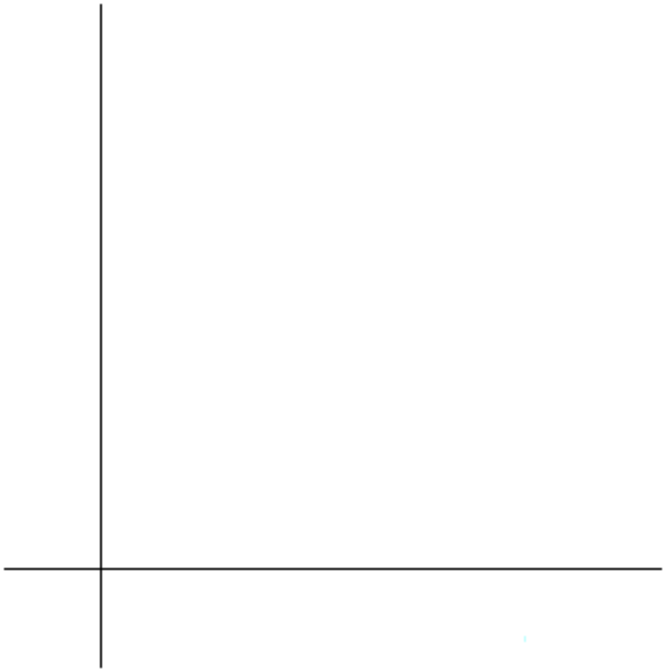
1. correlation coefficient = _____
2. p-value = _____
3. Is three significant correlation to use a regression line?

1. Calculate the following:

    - **a** =
    - **b** =
    - **c** =

- **d** equation: =

2. Obtain the graph using Minitab. Sketch the regression line below.



3. Supply an answer for the following senarios:

    - **a**For a textbook with 400 pages, predict the cost:
    - **b** For a textbook with 600 pages, predict the cost:

Discussion Questions

1. Answer each with 1-3 complete sentences.

    - **a**Does the line seem to fit the data? Why?
    - **b**What does the correlation imply about the relationship between the number of pages and the cost?

2. Are there any outliers? If so, which point(s) is an outlier?
3. Should the outlier, if it exists, be removed? Why or why not?

Collaborative Statistics: Group Project - Teegarden

Student Learning Objectives

- The student will identify a hypothesis testing problem in print.
- The student will conduct a survey to verify or dispute the results of the hypothesis test.
- The student will summarize the article, analysis, and conclusions in a report.

Instructions

As you complete each task below, check it off. Answer all questions in your summary. This project may be done in pairs or a group of three. Be sure to ensure that all the students participate equally in the work. This project is worth 20% of your final grade.

- **____Find an article** in a newspaper, magazine or on the internet which makes a claim about **ONE** population mean or **ONE** population proportion. The claim may be based upon a survey that the article was reporting on. Decide whether this claim is the null or alternate hypothesis.
- **____Copy or print out the article** and include a copy in your project, along with the source.
- **____State how you will collect your data.** (Convenience sampling is not acceptable.)
- **____Conduct your survey. You must have more than 50 responses in your sample.** When you hand in your final project, attach the tally sheet or the packet of questionnaires that you used to collect data. Your data must be real.
- **____State the statistics** that are a result of your data collection: sample size, sample mean, and sample standard deviation, OR sample size and number of successes.
- **____Make 2 copies of the appropriate solution sheet.**
- **____Record the hypothesis test** on the solution sheet, based on your experiment. **Do a DRAFT solution** first on one of the solution sheets and check it over carefully. Have a classmate check your solution to

see if it is done correctly. Make your decision using a 5% level of significance. Include the 95% confidence interval on the solution sheet.

- **____Create at least two different graphs to illustrate your data.** This may be a pie or bar chart or may be a histogram or box plot, depending on the nature of your data. Produce graphs that makes sense for your data and gives useful visual information about your data. Include an analysis of the graphs in your summary.
- **____Write your summary** (in complete sentences and paragraphs, with proper grammar and correct spelling) that describes the project. The summary **MUST** include:

    - **1**Brief discussion of the article, including the source.
    - **2**Statement of the claim made in the article (one of the hypotheses).
    - **3**Detailed description of how, where, and when you collected the data, including the sampling technique. Did you use cluster, stratified, systematic, or simple random sampling (using a random number generator)? As stated above, convenience sampling is not acceptable.
    - **4**Discuss the shape of your data and the relevant inforamition obtained from your graphs.
    - **5**Conclusion about the article claim in light of your hypothesis test. This is the conclusion of your hypothesis test, stated in words, in the context of the situation in your project in sentence form, as if you were writing this conclusion for a non-statistician.
    - **6**Sentence interpreting your confidence interval in the context of the situation in your project.

Assignment Checklist

Turn in the following typed (12 point) and stapled packet for your final project:

- **____Cover sheet** containing your name(s), class time, and the name of your study.
- **____Summary**, which includes all items listed on summary checklist.

- **____Solution sheet** neatly and completely filled out. The solution sheet does not need to be typed.
- **____Graphic representation of your data**, created following the guidelines discussed above. Include only graphs which are appropriate and useful.
- **____Raw data collected AND a table summarizing the sample data** (n, xbar and s; or x, n, and p', as appropriate for your hypotheses). The raw data does not need to be typed, but the summary does. Hand in the data as you collected it. (Either attach your tally sheet or an envelope containing your questionnaires.)